



《Spark编程基础（Python版）》

教材官网: <http://dblab.xmu.edu.cn/post/spark-python/>

温馨提示: 编辑幻灯片母版, 可以修改每页PPT的厦大校徽和底部文字

《Spark编程基础（Python版）》课程介绍

(PPT版本号: 2020年1月版)



扫一扫访问教材官网

林子雨

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn ➡

主页: <http://www.cs.xmu.edu.cn/linziyu>





提纲

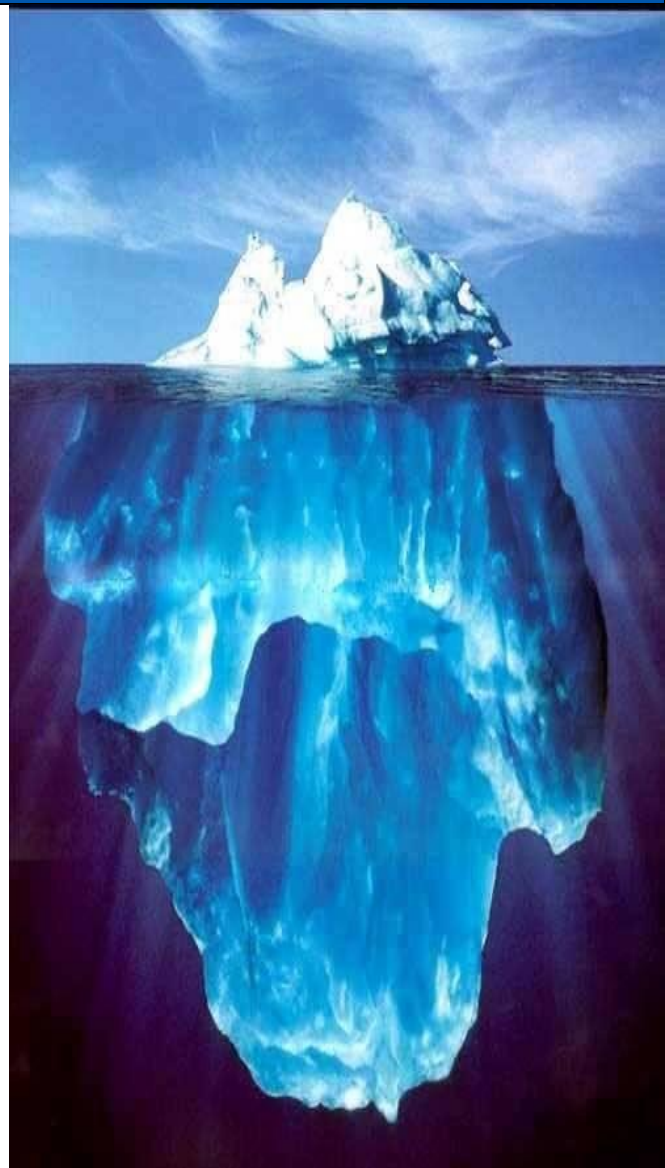
- 1.学习路线
- 2.先修课程
- 3.课程特色
- 4.教材介绍
- 5.内容提要
- 6.教学大纲
- 7.配套资源



高校大数据课程

公 共 服 务 平 台

百度搜索厦门大学数据库实验室网站访问平台





主讲教师



2018年国家精品在线开放课程

主讲教师：林子雨

中国高校首个“数字教师”提出者和建设者

2009年7月从事教师职业以来

累计**免费**网络发布超过**500万**字高价值教学和科研资料

网络浏览量超过**500万**次



数字教师LOGO



大数据学习路线图



大数据学习路线图访问地址: <http://dblab.xmu.edu.cn/post/10164/>



大数据学习路线图



子雨大数据

科普教育

纸质教材：《**大数据导论（通识课版）**》
视频：《**大数据导论（通识课版）**》MOOC视频

通识课



入门学习

纸质教材：《**大数据技术原理与应用**》
实验指导书：《**大数据基础编程、实验和案例教程**》
视频：《**大数据技术原理与应用**》MOOC视频
案例：大数据课程实验案例“网站用户购物行为分析”

专业课



进阶学习

纸质教材：《**Spark编程基础**》
在线教程：《**Spark入门教程（Scala版）**》
在线教程：《**Spark入门教程（Python版）**》
视频：《**Spark编程基础**》MOOC视频
案例1：淘宝双11数据分析与预测
案例2：Spark+Kafka构建实时分析Dashboard

专业课



实训案例

大数据实训案例教材
《**电信用户行为分析**》
《**基于协同过滤算法的电影推荐**》
《**互联网广告预测分析**》
《**微博用户情感分析**》

专业课



先修课程

在学习本课程之前，需要学生已经学习过如下课程

- 大数据技术导论
 - 比如林子雨编著《大数据技术原理与应用（第2版）》
 - 在线视频：<http://dbllab.xmu.edu.cn/post/bigdata-online-course/>
- 数据库系统原理
 - 了解关系数据库与SQL语句
- 编程
 - 比如，C++或者Java，了解面向对象编程基础知识



先修课程

建议在学习本课程之前，先学习大数据导论课程《大数据技术原理与应用》

大数据教材



1+1黄金组合
厦门大学林子雨编著

配套实验指导书





先修课程

建议在学习本课程之前，先学习大数据导论课程《大数据技术原理与应用》
荣获“**2018年国家精品在线开放课程**”

BIG DATA
打开大数据之门，畅游大数据世界

厦门大学
“大数据技术原理与应用”

报名参加 中国大学 MOOC 2017 年 11 月 6 日正式开课



厦门大学林子雨老师主讲《大数据技术原理与应用》课程视频
在线课程地址：<http://dblab.xmu.edu.cn/post/bigdata-online-course/>



课程特色



开启Spark学习大门
扫除学习障碍，铺平学习道路



课程教材

电子教材:

Spark入门教程(Python版)

<http://dblab.xmu.edu.cn/blog/1709-2/>



厦门大学林子雨



子雨大数据之Spark入门教程

披荆斩棘，在大数据丛林中开辟学习捷径



扫一扫访问在线教程

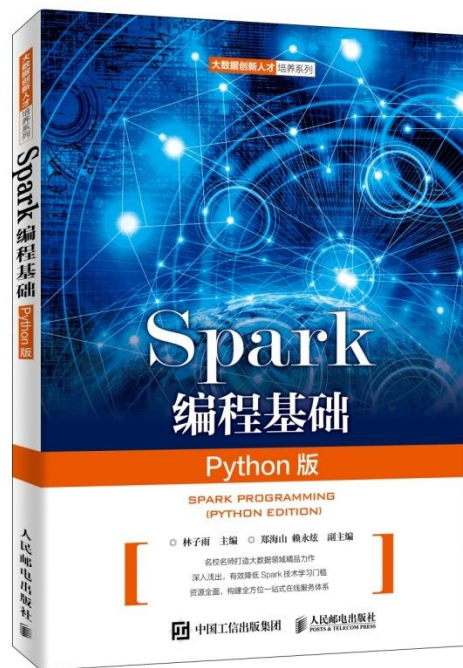


课程教材

林子雨，郑海山，赖永炫 编著 《Spark编程基础（Python版）》

教材官网：<http://dbllab.xmu.edu.cn/post/spark-python/>

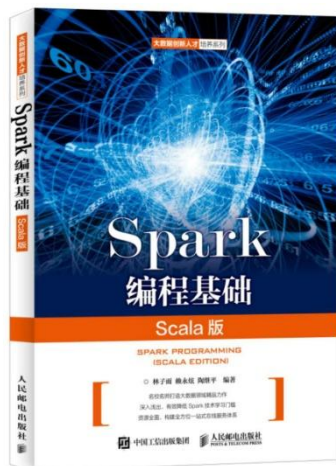
ISBN:978-7-115-52439-3 人民邮电出版社



本书以Python作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Structured Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。



参考书籍



《Spark编程基础（Scala版）》

厦门大学 林子雨，赖永炫，陶继平 编著

披荆斩棘，在大数据丛林中开辟学习捷径
填沟削坎，为快速学习Spark技术铺平道路
深入浅出，有效降低Spark技术学习门槛
资源全面，构建全方位一站式在线服务体系

人民邮电出版社出版发行，ISBN:978-7-115-48816-9
教材官网：<http://dblab.xmu.edu.cn/post/spark/>



本书以Scala作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Scala语言基础、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。



参考书籍



厦门大学林子雨



子雨大数据之Spark入门教程

披荆斩棘，在大数据丛林中开辟学习捷径



Spark在线教程地址： <http://dblab.xmu.edu.cn/blog/spark/>

如果没有购买《Spark编程基础》纸质教材，也可以使用Spark在线教程开展学习。Spark在线教程，比《Spark编程基础》纸质教材包含了更多的内容，并且同时提供了Scala版本和Python版本的两种Spark编程教程。在具体知识点方面，《Spark编程基础》纸质教材，要比Spark在线教程写得更加规范、详细和深入。二者各有千秋，可以结合使用。



课程内容提要

- **Spark**是当前热门的大数据处理技术
- 本课程重点介绍**Spark**的技术原理与编程方法
- **Spark**支持采用**Scala**、**Java**、**Python**和**R**语言进行编程，本课程采用**Python**语言编写**Spark**应用程序
- 课程内容包括**Spark**简介、设计与运行原理、**RDD**编程、**Spark SQL**、**Spark Streaming**、**Structured Streaming**、**Spark MLlib**等
- 通过本课程的学习，将帮助学生形成基础的**Spark**应用程序编程能力，为学生使用**Spark**技术解决实际科研问题和未来从事大数据相关工作奠定基础



教学大纲

章（或节）	主要内容	学时安排
第1章 大数据技术概述	大数据的基本概念、关键技术和代表性软件	2
第2章 Spark的设计与运行原理	Spark简介、运行架构、RDD的设计与运行原理、部署模式	3
第3章 Spark环境搭建和使用方法	Spark的基本安装方法，如何在spark-shell中运行代码以及如何开发Spark独立应用程序	3
第4章 RDD编程	RDD编程、键值对RDD、数据读写	4
第5章 Spark SQL	Spark SQL简介、DataFrame、读写Parquet、通过JDBC连接数据库、连接Hive读写数据	4
第6章 Spark Streaming	流计算简介、Spark Streaming简介、DStream操作	4
第7章 Structured Streaming	Structured Streaming简介、编程方法、输入源、输出操作、容错处理、迟到数据处理、查询的管理和监控	4
第8章 Spark MLlib	Spark MLlib简介、spark.mllib库、spark.ml库	4
综合案例	Spark应用程序案例	4
合计		32



课程配套教学资源

所有资料全部免费共享
支持电脑和手机浏览



高校大数据课程
公 共 服 务 平 台



高校大数据课程公共服务平台精华资源<http://dblab.xmu.edu.cn/post/8197/>



扫一扫访问平台主页



扫一扫观看3分钟FLASH动画宣传片



课程配套教学资源

课程实验

实验1-Linux系统的安装和常用命令

实验2-Spark和Hadoop的安装

实验3-RDD编程初级实践

实验4-Spark SQL编程初级实践

实验5-Spark Streaming编程初级实践

实验6-Structured Streaming编程实践

实验7-Spark机器学习库MLlib编程实践

可以访问教材官网获取以上课程实验

<http://dblab.xmu.edu.cn/post/spark-python/>



课程配套教学资源

Spark课程综合实验案例1：淘宝双11数据分析与预测

本案例涉及数据预处理、存储、查询和可视化分析等数据处理全流程所涉及的各种典型操作，涵盖Linux、MySQL、Hadoop、Hive、Sqoop、Eclipse、ECharts、Spark等系统和软件的安装和使用方法



案例访问地址：<http://dblab.xmu.edu.cn/post/8116/>



课程配套教学资源

课程实验案例2：Spark+Kafka构建实时分析Dashboard案例

由厦门大学数据库实验室团队开发，旨在满足全国高校大数据教学对实验案例的迫切需求。本案例涉及数据预处理、消息队列发送和接收消息、数据实时处理、数据实时推送和实时展示等数据处理全流程所涉及的各种典型操作，涵盖Linux、Spark、Kafka、Flask、Flask-SocketIO、Highcharts.js、sockert.io.js、PyCharm等系统和软件的安装和使用方法。案例适合高校（高职）大数据教学，可以作为学生学习大数据课程后的综合实践案例。



免费在线实验案例主页：<http://dblab.xmu.edu.cn/post/8274/>



附录A：主讲教师林子雨简介



主讲教师：林子雨

单位：厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn

个人网页: <http://dblab.xmu.edu.cn/post/linziyu>

数据库实验室网站: <http://dblab.xmu.edu.cn>



扫一扫访问个人主页

林子雨，男，1978年出生，博士（毕业于北京大学），现为厦门大学计算机科学系助理教授（讲师），曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革委员会副局长。中国计算机学会数据库专业委员会委员，中国计算机学会信息系统专业委员会委员。国内高校首个“数字教师”提出者和建设者，厦门大学数据库实验室负责人，厦门大学云计算与大数据研究中心主要建设者和骨干成员，2013年度和2017年度厦门大学教学类奖教金获得者，荣获2017年福建省精品在线开放课程、2018年厦门大学高等教育成果特等奖、2018年福建省高等教育教学成果二等奖、2018年国家精品在线开放课程。主要研究方向为数据库、数据仓库、数据挖掘、大数据、云计算和物联网，并以第一作者身份在《软件学报》《计算机学报》和《计算机研究与发展》等国家重点期刊以及国际学术会议上发表多篇学术论文。作为项目负责人主持的科研项目包括1项国家自然科学基金青年基金项目(No.61303004)、1项福建省自然科学基金项目(No.2013J05099)和1项中央高校基本科研业务费项目(No.2011121049)，主持的教改课题包括1项2016年福建省教改课题和1项2016年教育部产学协作育人项目，同时，作为课题负责人完成了国家发改委城市信息化重大课题、国家物联网重大应用示范工程区域试点泉州市工作方案、2015泉州市互联网经济调研等课题。中国高校首个“数字教师”提出者和建设者，2009年至今，“数字教师”大平台累计向网络免费发布超过500万字高价值的研究和教学资料，累计网络访问量超过500万次。打造了中国高校大数据教学知名品牌，编著出版了中国高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》，并成为京东、当当网等网店畅销书籍；建设了国内高校首个大数据课程公共服务平台，为教师教学和学生学习大数据课程提供全方位、一站式服务，年访问量超过200万次。





附录C：《大数据技术原理与应用》教材

《大数据技术原理与应用——概念、存储、处理、分析与应用（第2版）》，由厦门大学计算机科学系林子雨博士编著，是国内高校第一本系统介绍大数据知识的专业教材。人民邮电出版社 ISBN:978-7-115-44330-4 定价：49.80元



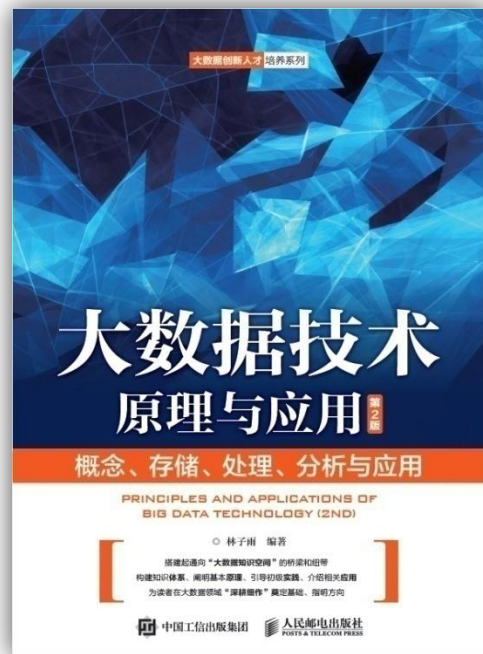
扫一扫访问教材官网

全书共有15章，系统地论述了大数据的基本概念、大数据处理架构Hadoop、分布式文件系统HDFS、分布式数据库HBase、NoSQL数据库、云数据库、分布式并行编程模型MapReduce、Spark、流计算、图计算、数据可视化以及大数据在互联网、生物医学和物流等各个领域的应用。在Hadoop、HDFS、HBase和MapReduce等重要章节，安排了入门级的实践操作，让读者更好地学习和掌握大数据关键技术。

本书可以作为高等院校计算机专业、信息管理等相关专业的大数据课程教材，也可供相关技术人员参考、学习、培训之用。

欢迎访问《大数据技术原理与应用——概念、存储、处理、分析与应用》教材官方网站：

<http://dbllab.xmu.edu.cn/post/bigdata>





附录D：《大数据基础编程、实验和案例教程》

本书是与《大数据技术原理与应用（第2版）》教材配套的唯一指定实验指导书

大数据教材



1+1黄金组合
厦门大学林子雨编著

配套实验指导书



- 步步引导，循序渐进，详尽的安装指南为顺利搭建大数据实验环境铺平道路
- 深入浅出，去粗取精，丰富的代码实例帮助快速掌握大数据基础编程方法
- 精心设计，巧妙融合，五套大数据实验题目促进理论与编程知识的消化和吸收
- 结合理论，联系实际，大数据课程综合实验案例精彩呈现大数据分析全流程

清华大学出版社 ISBN:978-7-302-47209-4 定价：59元

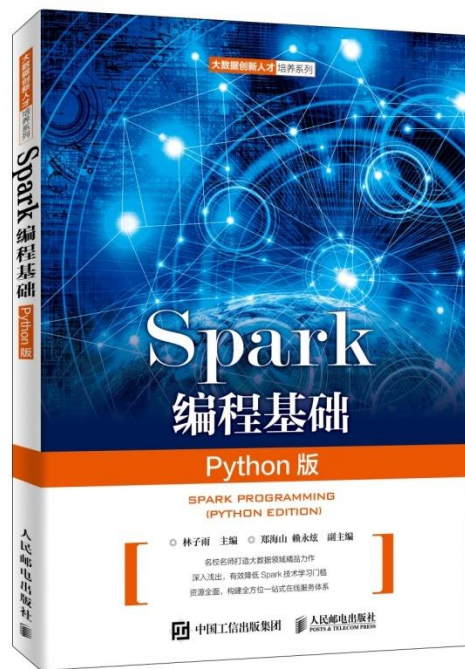


附录E：《Spark编程基础（Python版）》

林子雨，郑海山，赖永炫 编著 《**Spark编程基础（Python版）**》

教材官网：<http://dblab.xmu.edu.cn/post/spark-python/>

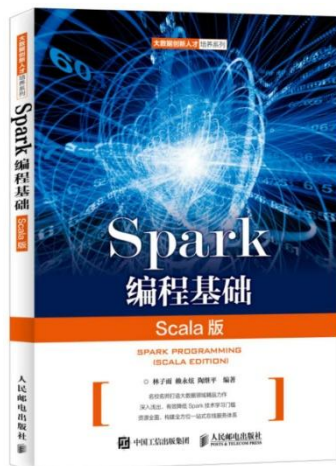
ISBN:978-7-115-52439-3 人民邮电出版社



本书以Python作为开发Spark应用程序的编程语言，系统介绍了Spark编程的基础知识。全书共8章，内容包括大数据技术概述、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Structured Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源，包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。



附录F: 《Spark编程基础 (Scala版)》



《Spark编程基础 (Scala版)》

厦门大学 林子雨, 赖永炫, 陶继平 编著

披荆斩棘, 在大数据丛林中开辟学习捷径
填沟削坎, 为快速学习Spark技术铺平道路
深入浅出, 有效降低Spark技术学习门槛
资源全面, 构建全方位一站式在线服务体系

人民邮电出版社出版发行, ISBN:978-7-115-48816-9
教材官网: <http://dbllab.xmu.edu.cn/post/spark/>



本书以Scala作为开发Spark应用程序的编程语言, 系统介绍了Spark编程的基础知识。全书共8章, 内容包括大数据技术概述、Scala语言基础、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作, 以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源, 包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。



附录G：高校大数据课程公共服务平台



高校大数据课程

公 共 服 务 平 台

<http://dblab.xmu.edu.cn/post/bigdata-teaching-platform/>



扫一扫访问平台主页



扫一扫观看3分钟FLASH动画宣传片

The background of the slide features a blue gradient with faint, light-blue silhouettes of groups of people. In the top left, a group of four people is holding hands. In the top center, a group of seven people is standing in a line. In the bottom left, two people are shown in profile, facing each other. On the right side, a person is standing with their hand on their chin, appearing to be in deep thought.

Thank You!

Department of Computer Science, Xiamen University, 2020